DOCUMENT RESUME

ED 226 731 IR 050 056

AUTHOR Long, Philip L.
TITLE Study and Preparation of Message Text Formats for NTAG.
SPONS AGENCY Council on Library Resources, Inc., Washington, D.C.; National Commission on Libraries and Information Science, Washington, D. C.; National Science Foundation, Washington, D.C.
PUB DATE 1 Feb 78
NOTE 39p.; For related document, see ED 185 976.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Computer Programs; Computers; Databases; *Indexes; Information Networks; Information Retrieval; Information Science; *Information Systems; Library Automation; *Library Networks; Library Technical Processes; *Online Systems; Search Strategies; *Standards
IDENTIFIERS *Computer Networking Research; *Message Text Formats

ABSTRACT

Prepared for consideration by the Network Technical Architecture Group (NTAG), a subcommittee of the Library of Congress (LC) Network Advisory Group, this document describes the study and development of text formats for the message delivery system of a nationwide library bibliographic network. Following a discussion of network considerations of the National Commission on Library and Information Science (NCLIS)/National Bureau of Standards (NBS) end to end computer protocol, an explanation is given of 13 command codes, 4 status codes, and 7 response codes which could be used to communicate between the many different types of computers in the proposed nationwide network. The key structure of search arguments is then identified for index types used by eight online bibliographic utilities or libraries with online systems, including: LC, the National Library of Medicine (NLM), OCLC, the University of Chicago, the Washington Library Network (WLN), Northwestern University, the New York Public Library, and BALLOTS. In order to facilitate cross-system comparison, the key structure of search arguments is also presented by type of file index. Author/title, CODEN, corporate/conference name, International Standard Book Number (ISBN), International Standard Serial Number (ISSN), LC classification number, personal name, subject, and title indexes are included. Finally, network standard search argument syntaxes are suggested for each type of index. (ESR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED226731

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

STUDY AND PREPARATION OF MESSAGE TEXT FORMATS FOR NTAG

by

Philip L. Long

IR050056

The materials incorporated in this report were developed with the financial support of the following:

National Science Foundation

National Commission on Libraries and Information Science

Council on Library Resources

However, the opinions, findings, conclusions, and recommendations expressed herein are those of the authors and do not necessarily reflect the views of the funding agencies identified above.

NCLIS/NBS END TO END PROTOCOL CONSIDERATIONS

The essential nature of the NCLIS/NBS end to end computer protocol is such that it allows a program operating in one machine to control all phases of the behavior of a receiving program which is resident in a remote machine. Peripheral I/O devices, be they card readers or CRT's, are not a consideration of the protocol. At each of a hierarchy of levels a specific form of interface is defined (see Figure 1). The consequence of this is that the precise nature of the data link control procedure and/or the procedure between a host computer facility and the facilities of a value-added carrier is completely immaterial to the NCLIS/NBS protocol as such. The protocol does assume, however, that the nature of the data link control procedure is such that transparent binary data may be transported from one computer to the other, so that facsimile or other noncharacter-type data may be transmitted in a record format yet to be established.

It is also presumed that each host computer facility participating in a national network will necessarily construct an interface programming system to effect a two-way interface for purposes of national networking. On the one hand, this interface programming (at the HOST level in Figure 1) must take care of the details of communicating via the NCLIS/NBS protocol and through such intermediate protocols which may be required at a particular site with the communications network itself. The other role of this programming is to "drive" such applications programs as exist or may be written in the future at a particular site, such

that this latter programming behaves according to the directives established through a network registry to enable the host facility computer to participate in the data exchange activities of the network.

There is no place within the NCLIS/NBS protocol for free text or word responses or commands. All communications, commands, and status and error replies are rigidly encoded as numbers in one or more header fields, as depicted below in Figure 2. Bibliographic data, brief entries, search arguments, etc., are placed in the data field of an NCLIS/NBS record. And such data are to be placed within a format devised for such data, e.g., Z39.2-1971 for bibliographic data. Data elements must be agreed upon and chosen for network brief bibliographic entries, and formats must be devised for search arguments. Indeed, the designers of that protocol contemplated that many such formats would need to be formulated as network implementation using the protocol takes place.

While the construction and testing of this particular software will not be required until an implementation of Phase 1.3 of the outlined activities of NTAG in this area are commissioned, substantial forethought in the construction of items for purposes of registration and implementation will be required prior to that time.

The consultant has identified for consideration of the NTAG Committee three areas requiring particular attention with regard to the NCLIS/NBS protocol. Experience gained in the early and intermediate stages of Phase 1.3 of the contemplated NTAG

activity will undoubtedly suggest refinements in these areas.

The areas (NCLIS/NBS terms) are those of commands, status responses, and response codes, the last being transmitted by the computer facility which is providing a service (the target facility), such as bibliographic search, to the host computer facility which has originated the request, in the event that the search has failed.

IA. The consultant has identified the following COMMAND CODES (20 through 28) for search codes for network NORMALIZED search arguments (MSGTYPE's for FID 2 of the NCLIS/NBS protocol -- see Figure 2), and their meanings for queries:

*20. Perform bibliographic search - non-standard argument. This command instructs the target facility to perform a search against a target facility index using a search argument that is non-standard on the network. The index or indexes, type and format of the argument are subject to bilateral agreement between host and target facility for all exceptions to those provided below. Use of this command code is not recommended.

---

* The choice of numbers given here is arbitrary, and intended to avoid use of lesser valued numbers which appear in the NCLIS/NBS document.

21. Perform bibliographic search - name. This command instructs the target facility to perform a search against the target facility name index for bibliographic records which match the criterion established by the given search key string.

22. Perform bibliographic search - title. This command instructs the target facility to perform a search against its title index.

23. Perform bibliographic search - subject. This command instructs the target facility to search for bibliographic records matching the given subject search key string criteria.

24. Perform bibliographic search - author/title. This command instructs the target facility computer to perform a bibliographic search using a precoordinated author and title key.

25. Perform bibliographic search - CODEN. This command instructs the target facility computer to perform a bibliographic search using a given CODEN as the search criterion.

26. Perform bibliographic search - ISBN. command instructs the target facility computer to perform a bibliographic search using the International Standard Book number which accompanies the search command.

27. Perform bibliographic search - ISSN. In like fashion, this command instructs the target facility computer to perform a bibliographic search using the International Standard Serial number. (Strictly speaking, since the scope of this study is restricted to monographic searches, this command is not to be considered part of the regular command structure. However, in consideration of the body of literature and category of monographic series, the command has nonetheless been included.)

28. Perform bibliographic search - LCCN. This command instructs the target facility computer to perform a bibliographic search using the Library of Congress card number as the search criterion.

IB. The following command codes (29 and 30) have been identified for use in conjunction with search codes for network NORMALIZED search arguments 20 through 28 above (MSGTYPE's for FID 2 of the NCLIS/NBS protocol -- see Figure 2) for determining the status of

a search or aborting an ongoing search:

29. Transmit status of search. This command instructs the target facility computer to reply to the host computer requesting the service with a status message indicating either the reason for a particular task failure in a search or the nature of current activity within the target facility. It must be noted that all computer systems may not be capable of accurately reporting the nature of ongoing activity. Thus this area of command structure and the accompanying status reply will require particularly careful consideration in the implementation phase.

30. Abort the search. This command instructs the target facility to immediately cease action on an ongoing bibliographic search. This command, like the immediately preceding command, will require especially careful consideration in the implementation phase.

IC. Bibliographic record responses are elicited by command codes 31 and 32 within the NCLIS/NBS protocol (MSGTYPE's for FID 2) following a query (command codes 20 through 28):

31. Transmit the next (or first) block of brief bibliographic records. This command instructs the target facility computer to reply with the next record or group of records which have satisfied a particular search key criterion. The brief records are to be in Z39.2 format.

32. Transmit the next (or first) block of full bibliographic records. This command is similar to the immediately preceding command, except that the desired reply will consist of ANSI Z39.2-formatted (e.g., MARC-II) records.

II. The consultant hasd identified the following list of STATUS-TYPE codes (STA in FID 8 of the NCLIS/NBS prot[illegible] -- see Figure 2) to be emitted by the target facility under the NCLIS/NBS protocol in aid of bibliographic searching on the remote target facility computer system in response to a type 29 command (transmit status of search):

21. Search is awaiting start. This response indicates that the request for search has been received at the target facility, but for one or another reason, has not yet actually been placed into an execution state.

22. Search is being conducted. This status reply indicates that requested search is being processed at the target facility.

23. Brief bibliographic record replies (still) await transmission. This status reply indicates that the given bibliographic search has been successfully concluded, or that a logical break point has occurred in that search. It is contemplated by the consultant that the response from the originating computer facility to the receipt of this status would be either a type 31 or a type 32 command to elicit transmission of the stored replies to the query.

24. Full bibliographic record replies (still) await transmission. This status response is similar to status type 23, the exception being the implication that full bibliographic records await transmission. The consultant contemplates that the two computer systems would enter a conversational mode of gathering one or more blocks of records by interspersed use of type 23 and type 24 status responses as necessary coupled with type 31 or type 32 commands.

III. In the event that the target facility computer system is unable to satisfy a request for bibliographic record search, the target facility system is expected to respond to the originating host with an NCLIS/NBS message containing the response code to explain the reason for this situation or error condition. The consultant has identified the following RESPONSE TYPE CODES (also STA in FID 8 of the NCLIS/NBS protocol):

41. Ill-formed commnad. In the event that the target facility determines that the syntax of a command it has received is improperly formatted, this response code is to be transmitted in reply.

42. Unrecognized command. In the event the target facility computer programming detects a properly formatted command message but does not find the command code in its repertoire of registered commands, this code is transmitted in response.

43. Unsupported command. In the event the target facility computer systems receive a properly formatted command which exists within the national registry system but which calls for an action which is not provided by that particular target facility, this response is transmitted to the originating computer system.

44. Price limit exceeded. The NCLIS/NBS protocol provides (among others) the ability to cause termination of an ongoing search automatically provided a resource limit previously established in an interchange between the originating and target facilities has been exceeded. The resource limit in this case is a monetary sum limit. This situation constitutes an abort condition. Whether or not any intermediate results stored at the target facility are to remain available to the originating system, and if so for what time period, is a topic which remains to be resolved by discussion among the members of NTAG.

45. Permanent target system limit exceeded. This response code indicates to the originating system that some resource at the target facility, which could be required to satisfy the search, is not available. An example of this might be a sufficiently large sized system buffer. This condition constitutes a permanent abort, such that the originating facility is thus notified that reattempting the search will be of no avail. Whether or not any intermediate results are to be retained at the target facility and/or whether or not the target facility may legitimately charge

the originating facility for having attempted the search remains a matter for discussion among members of the NTAG Committee.

46. Temporary target facility system limit-exceeded. This response code indicates that the non-permanent system resource required for successful completion of the search was not available at the time of the execution of the search in the target facility system. One example of such a resource might be available scratch disc storage. Whether or not a charge is to be made for undertaking this aborted search is a matter for discussion similar to the previous error code. This important distinction between this type response code and the previous (type 45) response code, is that, in principle, the search may be successfully completed should the originating system repeat the request at a later time.

47. No record in target facility system matches search criteria. This response code indicates that the target facility data base contains no record which satisfies the search criterion as given by the key string which accompanied the search command. This response code does not indicate that no record exists in the target system which may satisfy a

users search criterion. This latter hypothesis can only be tested by transmittal of a following search type command using a different key.

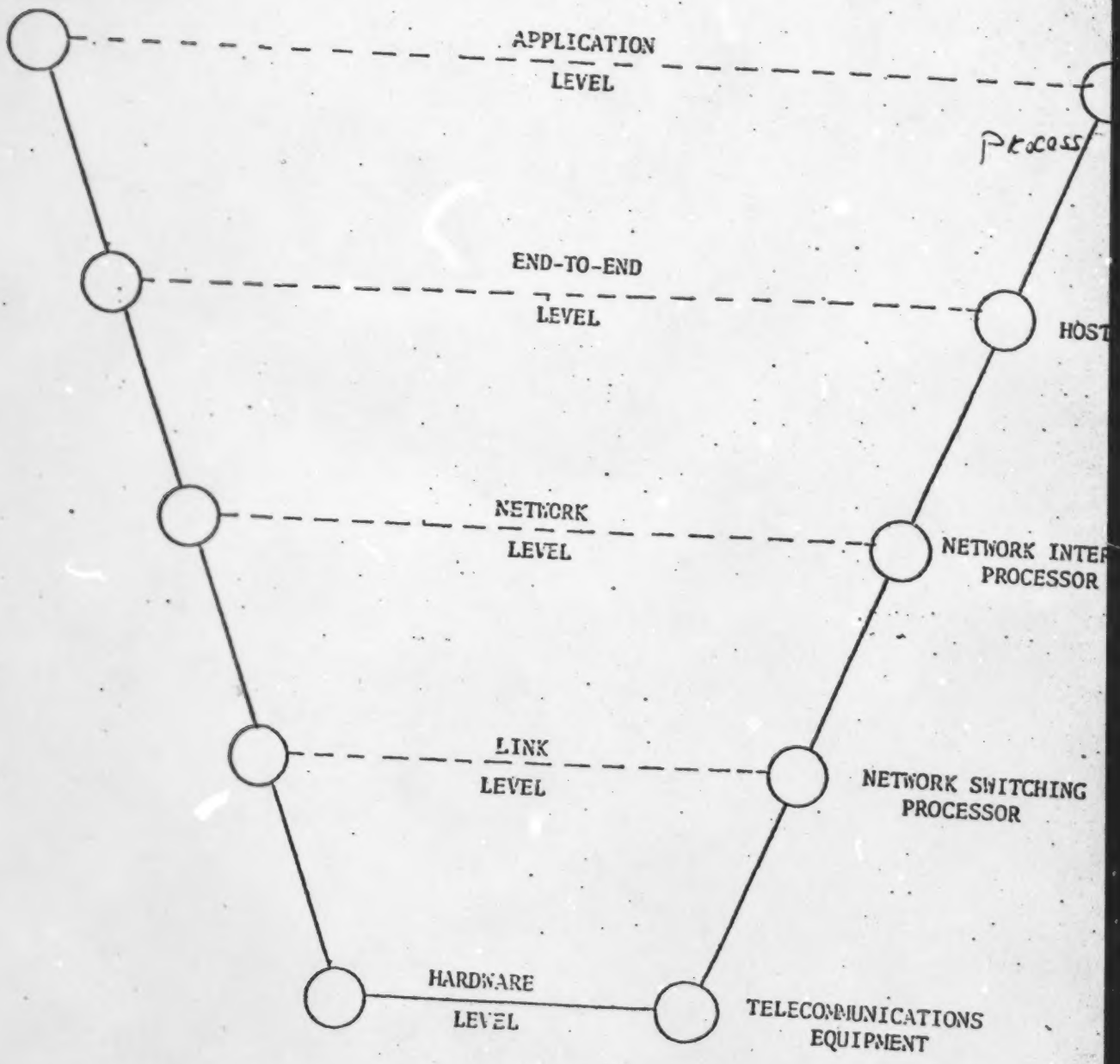

Figure 1
(Based on NCLIS/NBS Task Force on Computer Network Protocol)

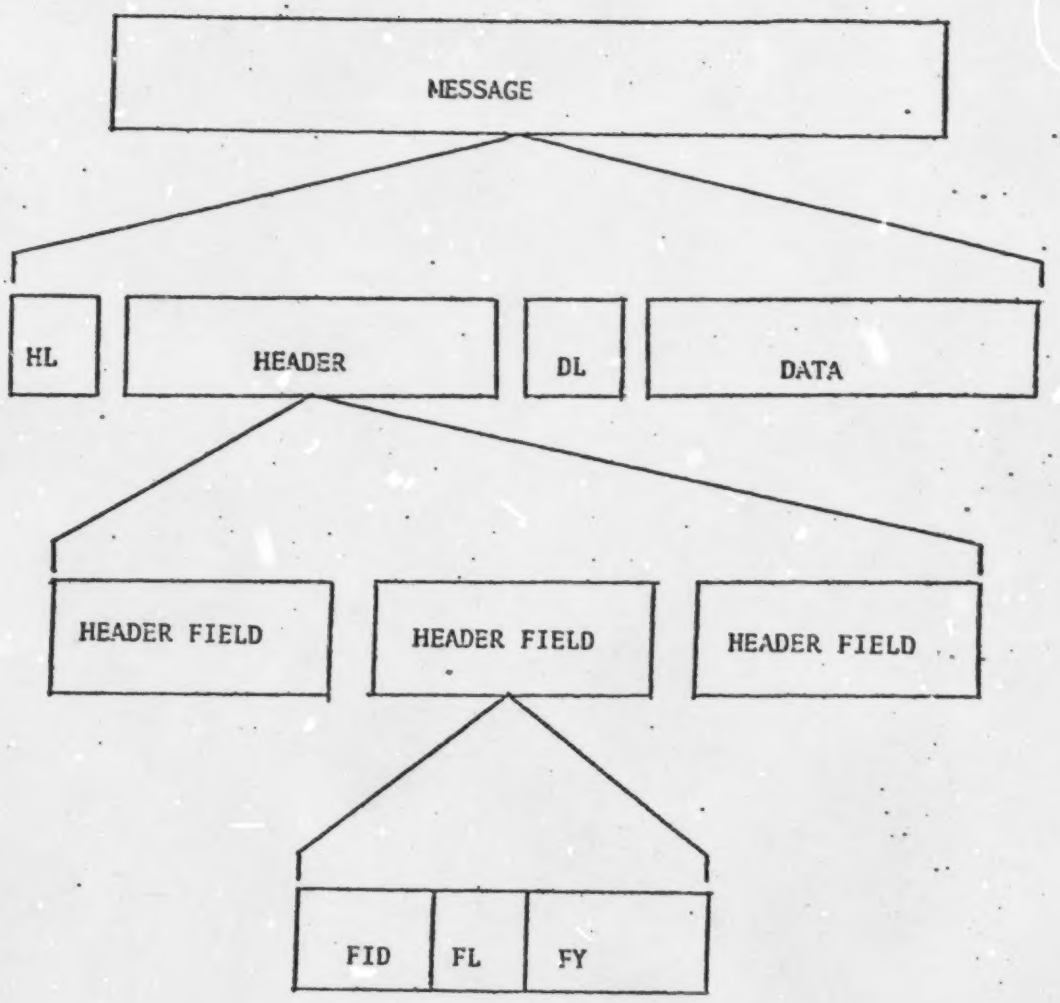

Figure 2

(Based on NCLIS/NBS Task Force on Computer Network Protocol)

NTAG STUDY

SEARCH KEY CONSIDERATIONS

The following data represents the pertinent search arguments in use at each institution, exclusive of explicit CRT-generated commands. This type of data is to be placed in the data area of NCLIS/NBS search command messages (see Figure 2). As mentioned above, text as currently found on screen displays is meaningless in the NCLIS/NBS protocol. Hence, only pertinent logical data are presented. Appropriate format and syntactic rules for transmittal of this data must be formulated for use in phase 1.3 of the NTAG plan. The magnitude of such an undertaking is quite large.

The representation used below to describe search arguments is as follows:

⟨ ⟩	indicates a string
[ ]	indicates optional data
w	indicates a word
C	indicates an alphanumeric character
a	indicates an alphabetic character
n	indicates a numeric character
#	indicates implicitor explicit truncation special characters are as seen

T for Northwestern Name/Title and Title/Name key structures only

superscripts indicate mandatory length when they occur alone and indicate maximum length when they occur with a subscript

subscripts indicate minimum length

Example: $[a_1^3]n^2\text{-}n_1^6$

represents a string with (optionally) one to three leading alphabetic characters, followed by exactly two numbers, followed by a hyphen, followed by one to six numbers. This is a typical format for an LC card number key string.

Since this study is directed to bibliographic searching (see section 2A, "TASK DEFINITION", of the CCNBC statement of work), non-bibliographic search methodologies (e.g., local accession number) are not presented below. In order to present an overview of similarities in local on-line search facilities, search argument syntax is presented; detail on search strategy is omitted. Lastly, since the protocol is computer-to-computer in orientation, terminal-operator oriented textual

messages, which are both highly ideosyncratic and subject to frequent alteration, are not presented.

Virtually all systems studied have both specific (e.g., ISBN) and ambiguous (e.g., title) access points. Ambiguous arguments may lead to display of a single record (hopefully the desired one), or to a list of brief entries from which a full record or list of full records may be obtained. Precisely which data elements in a network environment should appear in a brief entry is yet to be settled upon. Certainly these entries should be encoded in Z39.2 format. System by system argument syntax follows:

LC	Index type	Key structure
	Author/Title	c_1^3,c_1^3
	ISBN	n_9^{10}
	LCCN	$[a_1^3]n^2\text{-}n_1^6$
	Personal Name	c_1^n,c_1^m
	or	c^5,c
	Title	$c_1^3[,C][,C][,C]$

NLM*

Corporate of Conf. Name	$\langle w_1^n \rangle$
ISBN	$n - n_1^x - n_1^y - n^1$ (BN), $x+y=8$
ISSN	$n^4 - n^4$ (IS)
LCCN	$\langle n^2 - n_1^6 \rangle$
Main Heading (MeSH)	$\langle w_1^n \rangle / \langle w_1^n \rangle$
Name (corp) as subject	see corp. name
Name (pers) as subject	see personal name
Personal Name	w,w[#]
Corp. name series title	$\langle C_1^{39}[\#] \rangle$
Pers. name series title	$\langle C_1^{39}[\#] \rangle$
Series title	$\langle C_1^{39}[\#] \rangle$
Title	C^3[/C][/C][/C]

* Since all on-line NLM systems are based on a most sophisticated and extremely powerful, interactive information storage and retrieval system approach to data retrieval in a largely controlled - vocabulary environment that internally bears little resemblance to the other systems studied, only access points resembling those in use in primarily bibliographic-oriented systems are excerpted for presentation here.

OCLC

Index type	Key structure
Author	$c_1^4, [c_1^3], [c]$
Author/Title	$[c_1^4], c_1^4$
CODEN	a_5^6
ISBN	n^{10}
ISSN	$n^4 - n^4$
LCCN	$[a_1^3]n_1^2 - n_1^6$
Title	$c_1^3, [c_1^2], [c_1^2], [c]$

University of Chicago	Index type	Key structure
	Author/Title	$'C^3\ C_1^3'$
	ISBN	n^{10}
	ISSN	n^8
	LCCN	$[a_1^3]n^2\ -n_1^6$
	Title	$'C^3C\ C\ C'$

WLN	Index type	Key structure
	AUTHOR	$\langle(w[\#]_1^9\rangle$
	ISBN	n_1^{10}
	ISSN	$n^4 - n^4$
	LCCN	$[a_1^3]n^2[-]n_1^6$
	Subject	$\langle(w[\#])_1^9\rangle$
	Title	$\langle(w[\#])_1^9\rangle$

Northwestern/on-line monograph records

Index type	Key structure
Name/Title	$T_{mn}T_{sn}[D^4]$
Title/Name	$T_{mt}T_{st}[D^4]$

$D^4=n^4$ = MARC "date 1" - 008 field

$T_{mn}=T_{st}$ = personal or corporate entry

$= C_1^i[C_1^i]$ where $i \leq 70$, $j \leq 68-i$

$T_{st}=T_{mt}= C_1^i[C_1^j]$ where $i \leq 70$, $j \leq 68-i$

RLG/New York Public Library

Index type	Key structure
Author	$\langle c_1^{80}[\#] \rangle$
ISBN	n^{10}
LCCN	$n^2\text{-}n_1^6$
Title	$\langle c_1^{80}[\#] \rangle$

BALLOTS/full face

Index type	Key structure
Author (personal name)	$(C_3^n[\#])_1^n \quad w[\#][,C_1^n]([C_1^n])_1^m$
	or $\quad C_1^n[\#]C_1^n[\#]([C_1^n])_1^m$
LCCN	n_3^8
Name, corporate/conf.	$\langle(w[\#])_1^n\rangle$
Subject (topical, geographic)	$\langle C_3^n[\#]\{[\#]C_1^n[\#]\}\rangle$
Title	$\langle(w[\#])_1^n\rangle$

SEARCH ARGUMENT MATRIX

In order to facilitate cross-system comparison of search arguments, the foregoing information has been reorganized by type of index file. One can readily see that, at this time, network-canonical syntax access by CODEN and/or by corporate or conference name is moot.

Index type: Author/Title

System:	Argument Syntax
LC	C_1^3, C_1^3
NLM	None--Boolean coordination
OCLC	$[C_1^4], C_1^4$
U. of Chicago	$'C^3C_1^{3'}$
WLN	None
Northwestern (name/title)	$T_{mn}T_{sn}[D^4]$
NYPL	None
BALLOTS	None--Boolean coordination

Index type:	CODEN
LC	None
NLM	None
OCLC	a^{6}_{5}
U. of Chicago	None
WLN	None
Northwestern	None
NYPL	None
BALLOTS	None

Index type:	Corporate or Conference Name
System:	Argument Syntax
LC	None
NLM	$\langle w_1^n \rangle$
OCLC	None
U. of Chicago	None
WLN	None
Northwestern	None
NYPL	None
BALLOTS	$\langle (w[\#])_1^n \rangle$

Index type:	ISBN
System:	Argument Syntax
LC	n_9^{10}
NLM	$n - n_1^x - n_1^y - n\ (BN), x+y=8$
OCLC	n^{10}
U. of Chicago	n^{10}
WLN	n_1^{10}
Northwestern	None
NYPL	n^{10}
BALLOTS	None

Index type:	ISSN
System:	Argument Syntax
LC	None
NLM	$n^4 - n^4$ (IS)
OCLC	$n^4 - n^4$
U. of Chicago	n^8
WLN	$n^4 - n^4$
Northwestern	None
NYPL	None
BALLOTS	None

Index type:	LCCN
System:	Argument Syntax
LC	$[a_1^3]n^2 - n_1^6$
NLM	$\langle n^2 - n_1^6 \rangle$
OCLC	$[a_1^3]n_1^2\text{-}n_1^6$
U. of Chicago	$[a_1^3]n^2 - n_1^6$
WLN	$[a_1^3]n^2[\text{-}]n_1^6$
Northwestern	None
NYPL	$n^2 - n_1^6$
BALLOTS	n_3^8

Index type:	Personal Name
System:	Argument Syntax
LC	C_1^n, C_1^m or C^5, C
NLM	w,w[#]
OCLC	$C_1^4, [C_1^3], [C]$
U. of Chicago	None
WLN	$\langle (w[\#])_1^9 \rangle$
Northwestern	None
NYPL	$\langle C_1^{80}[\#] \rangle$
BALLOTS	$w[\#][,C_1^n]([C_1^n])_1^m$
or	$C_1^n[\#]\ C_1^n[\#]([C_1^n])_1^m$

Index type:		Subject
System:		Argument Syntax
LC		None
NLM	MeSH	$\langle w_1^n \rangle / \langle w_1^n \rangle$
	pers. or corp. name	$\langle w_1^n \rangle$
OCLC		None
U. Chicago		None
WLN		$\langle (w[\#])_1^9 \rangle$
Northwestern		None
NYPL		None
BALLOTS		$\langle C_3^n[\#] \left[[\#] C_1^n[\#]\right] \rangle$

Index type:		Title
System:		Argument Syntax
LC		$C_1^3[,C][,C][,C]$
NLM		$C^3[/C][/C][/C]$,
	for series	$\langle C_1^{39}[\#]\rangle$ (includ. pers. & corp. name)
OCLC		$C_1^3,[C_1^2],[C_1^2],[C]$
U. of Chicago		'C³ C C C'
WLN		$\langle(w[\#])_1^9\rangle$
Northwestern (title/Name)		$T_{mt}T_{st}[D^4]$
NYPL		$\langle C_1^{80}[\#]\rangle$
BALLOTS		$\langle(w[\#])_1^n\rangle$

SUGGESTED NETWORK NORMALIZED SEARCH ARGUMENT*

The consultant suggests the following as network standard search argument syntaxes for NTAG's consideration. Bibliographic records in reply are to be in LC Z39.2 format; tags for brief records must be chosen.

Index type	Suggested Standard Syntax
Author/title	$C_1^3\ C_1^3$
CODEN	a_5^6
Corporate or Conference Name	$\langle w_1^n \rangle$
ISBN	n^{10}
ISSN	n^8
LCCN	$[a_1^3]n^2\ n_1^6$
Personal Name	$\langle (w[\#])_1^9 \rangle$
Subject	$\langle (w[\#])_1^9 \rangle$
Title	$\langle (w[\#])_1^9 \rangle$

---

* These data are to be placed in the data field of the NCLIS/NBS message carrying a search command. Formats for these data are yet to be defined. N.B.: Status and error responses, and non-search commands have no data fields.